

# On stochastically sampling color configurations

Joshua Isaacson<sup>1,\*</sup> and Stefan Prestel<sup>1,†</sup>

<sup>1</sup>*Theoretical Physics Department, Fermi National Accelerator Laboratory, P.O. Box 500, Batavia, IL 60510, USA.*

Parton shower algorithms are key components of theoretical predictions for high-energy collider physics. Work towards more accurate parton shower algorithms is thus pursued along many different avenues. The systematic treatment of subleading color corrections in parton shower algorithms is however technically challenging and remains elusive. In this article, we present an efficient and numerically stable algorithm to sample color configurations at fixed  $N_C = 3$ , using the correct color factor including subleading corrections with a parton shower. The algorithm is implemented as stand-alone program that can be interfaced to the PYTHIA event generator. Preliminary comparisons to LEP data are presented.

## I. INTRODUCTION

High energy physics event generators are tools that enable comparisons of theory calculations to measurements at high-energy collider experiments [1]. This is achieved by simulating scattering events in as much detail as possible. Event generators are thus heavily used in analysis prototyping, to compare precision Standard-Model (SM) measurements to theory, or to derive indirect bounds on new phenomena. This is only possible if event generators incorporate precise calculations in perturbative QCD. Dedicated efforts over the last decade have ensured that scattering events with well-separated jets are described with high accuracy. The accuracy of the modeling of jet structure and evolution is much less understood. Jet evolution can be generated by parton showers [2] – programs that are based on solving the leading order DGLAP evolution equations [3] by explicitly constructing states with multiple branchings. Several approximations are then made to allow an iterative, probabilistic interpretation of these evolution equations. Precision SM measurements or indirect limit setting using jet substructure will mean that some if not all of the parton shower approximations will have to be reevaluated and removed.

This article describes an algorithm to overcome the  $N_C \rightarrow \infty$  limit (also known as the Leading Color (LC) approximation) typically employed by parton shower programs. In the past few years, there has been a resurgence in the desire to understand the formalism and requirements for full-color parton showers [4–9]. The goal of the current work is to present a realistic algorithm to efficiently and consistently implement a means to sample, according to single-parton insertion operators, arbitrary color configurations defined with  $N_C = 3$  within a parton shower. We introduce a new parton shower algorithm that strictly employs  $N_C = 3$  and can be matched onto non-perturbative hadronization models, which does not experience factorial growth in the time to process each subsequent emission that was previously observed [4]. The manuscript first introduces some basic parton shower concepts in Sec. II before describing the new  $N_C = 3$  parton shower algorithm in Sec. III. This is followed by results and comparisons to experimental measurements in Sec. IV, before giving conclusions and an outlook in Sec. V.

## II. PARTON SHOWER BASICS

Parton showers are a crucial component of event simulations, as they link low-multiplicity hard scatterings calculations to realistic jet observables by explicitly simulating perturbative jet formation and evolution. This is achieved by solving the leading-order DGLAP-like renormalization group equations

$$\frac{d f_a(x, t)}{d \ln t} = \sum_{b=q,g} \int_0^1 \frac{dz}{z} \frac{\alpha_s}{2\pi} [P_{ab}(z)]_+ f_b\left(\frac{x}{z}, t\right) \quad (1)$$

with Markovian Monte-Carlo algorithms, i.e. by interpreting the evolution kernels  $P_{ab}$  as probabilities with which to distribute real emissions. The virtual corrections defined at the kinematic endpoints (defined by the  $+$  prescription) are included by enforcing probability conservation when proposing state changes from parton branching. As a direct

\* isaacson@fnal.gov

† sprestel@fnal.gov

consequence, parton showers commonly generate only those virtual momentum and color exchanges that can be obtained from real-emission configurations upon integration<sup>1</sup>.

Modern parton showers extend this probabilistic evolution picture to soft radiation by including coherent emissions from color dipoles. For the simplest case of a color-anticolor connection, the soft limit of single-gluon emission can be obtained by ensuring that the sum of evolution kernels recovers eikonal factors, either differentially or at the integrated level, and by using the  $N_C = 3$  value of  $C_F = \frac{N_C^2 - 1}{2N_C} = \frac{4}{3}$ . Parton showers that recover eikonal factors differentially are typically based on a dipole picture [2, 11]. The combined soft-collinear evolution of dipole showers is governed by new dipole evolution kernels. These can be obtained by the matrix element factorization formula in the soft and collinear limits [12]

$$|\mathcal{M}_{n+1}|^2 \simeq - \sum_{\tilde{i}\tilde{j}, \tilde{k} \neq \tilde{i}\tilde{j}}^n \langle \mathcal{M}_n | \frac{\mathbf{T}_{\tilde{k}} \cdot \mathbf{T}_{\tilde{i}\tilde{j}}}{\mathbf{T}_{\tilde{i}\tilde{j}}^2} \mathcal{V}_{\tilde{i}\tilde{j}, \tilde{k}} | \mathcal{M}_n \rangle, \quad (2)$$

where parton  $\tilde{i}\tilde{j}$  is the parton undergoing the branching,  $\tilde{k}$  is a spectator, and  $\mathbf{T}_n$  is the color matrix element for parton  $n$ , and the sum over  $\tilde{i}\tilde{j}, \tilde{k}$  extends to all colored partons. This is then further simplified in the  $N_C \rightarrow \infty$  limit, and by discarding spin correlations between the hard matrix element and the branching, to

$$|\mathcal{M}_{n+1}|^2 \approx \sum_{\tilde{i}\tilde{j}, \tilde{k} \in \text{LC}}^n \langle \mathcal{M}_n | \mathcal{M}_n \rangle \mathcal{V}_{\tilde{i}\tilde{j}, \tilde{k}}. \quad (3)$$

where the sum runs only over partons connected in the  $N_C \rightarrow \infty$  limit. The dipole kernels  $\mathcal{V}_{\tilde{i}\tilde{j}, \tilde{k}}$  are often constructed to only contain two-particle poles in the collinear limits. This article will assume the construction of Catani and Seymour [12] in the reorganization of [12, 13], although the exact form of the kernels is not of immediate importance for the following.

The combined soft/collinear evolution equation is solved by iteratively proposing, accepting or rejecting state changes according to the parton branching kernels  $\mathcal{V}_{\tilde{i}\tilde{j}, \tilde{k}}$ . The resummation of large logarithmic enhancements into Sudakov form factors (i.e. no-emission or “parton survival” probabilities) with this procedure requires that state changes are probed in an ordered sequence. The ordering variable should isolate the soft and collinear regions in (at least) the single-emission phase space, and thus defines how phase space points are sampled and the exponentiation properties of the parton shower. The typical choices of using angle [14] or hardness-related variables like virtuality [2, 14] or transverse momentum [2, 11, 12, 14] are related to the on-shell propagator singularities induced by emissions. Transverse momenta measured with respect to the axes of a color dipole have favorable qualities in the soft limit [2, 15] and to define consistent loop integration boundaries [16]. This article uses the soft transverse momentum-ordered final-state shower as outlined in [13] and implemented in PYTHON in [17] as starting point.

Hardness-ordered parton showers also allow the use of a single scale value to both avoid the Landau pole in  $\alpha_s$  and to transition to phenomenological non-perturbative hadronization models. Such hadronization models typically employ a leading color QCD picture to assign the starting conditions for the parton-to-hadron conversion process. For a simple infrared-safe treatment of soft and collinear gluons in setting up the starting conditions for hadronization, we will use the leading color Lund string model [18] as implemented in the PYTHIA 8 event generator [19, 20]. The subtleties related to this phenomenological step are discussed in Sec. III D.

The algorithm described in this article remains within the paradigm of (weighted) unitary parton shower evolution [21–23]. Thus, the algorithm does not discuss the inclusion of Glauber gluons [9, 16, 24–26], that decouple the color structure of virtual corrections from that of real-emission diagrams. It is not obvious that the impact of Glauber gluons can be assessed without also including other universal virtual corrections, e.g. related to analytic continuation or remainders of  $d$ -dimensional loop integration in certain regularization schemes. It can be conjectured that such purely virtual corrections could be included by a “swing” mechanism [27]. Due to these difficulties, we postpone the inclusion of Glauber gluons to a future publication.

### III. PARTON SHOWER ALGORITHM AT FIXED $N_C$

This section describes a stochastic algorithm to systematically sample subleading color configurations in QCD. This relies on standard parton shower techniques, and will thus be referred to as fixed color (FC) parton shower. The main

---

<sup>1</sup> Developments in [10] discuss the extension of parton showers to include virtual corrections without direct real-virtual correspondence.

	$\lambda_i$	$\bar{\lambda}_i$
Quark	$\sqrt{T_R}$	0
Antiquark	0	$\sqrt{T_R}$
Gluon	$\sqrt{T_R}$	$\sqrt{T_R}$

TABLE I. Color variables in the color operators for quarks, antiquarks, and gluons.

goal is then to ensure the color correlations and interferences between different color structures encoded in Eq. 2 are retained by the FC parton shower algorithm. This idea has previously also been advocated in [4]. Here, we provide an algorithm that allows to handle an arbitrary number of emissions with less than factorial growth in complexity. This is realized by directly sampling color configurations in the color flow basis [28] by introducing color-generator-specific dipole splitting kernels, which allow to evolve from one definite color configuration to another definite color configuration. The full color space is sampled stochastically. This section first reviews the relevant benefits of splitting functions in the color flow basis (Sec. III A), before discussing the implementation as a FC parton shower (Secs. III B and III C), and describing the interface to leading color evolution and hadronization in Sec. III D.

### A. Splitting Kernels in the Color Flow Basis

The gluon propagator in QCD is proportional to

$$\langle (\mathcal{A}_\mu)_{j_1}^{i_1} (\mathcal{A}_\mu)_{j_2}^{i_2} \rangle \propto \delta_{j_2}^{i_1} \delta_{j_1}^{i_2} - \frac{1}{N_C} \delta_{j_1}^{i_1} \delta_{j_2}^{i_2}. \quad (4)$$

Thus, the gluon can thus be treated as two distinct contributions, the “nonet” gluon (the terms proportional to  $\delta_{j_2}^{i_1} \delta_{j_1}^{i_2}$ ) and the “singlet” gluon (the terms proportional to  $\delta_{j_1}^{i_1} \delta_{j_2}^{i_2}$ ). This interpretation is the backbone of the color flow basis [28]. Note that the singlet contribution to the gluon propagator is suppressed by a factor of  $1/N_C$ .

The color flow basis can conveniently be represented by a set of basis tensors represented with color indices ( $c_i$ ) and anticolor indices ( $\bar{c}_i$ ) [9]. In this representation, the tensors are represented by a color-anticolor pair, which denotes the flow of color from one leg to another, and can be expressed in terms of Kronecker  $\delta$ 's,

$$|\sigma\rangle = \left| \begin{array}{cccc} 1 & 2 & \dots & n \\ \bar{\sigma}(1) & \bar{\sigma}(2) & \dots & \bar{\sigma}(n) \end{array} \right\rangle = \delta_{c_{\sigma(1)}}^{c_1} \delta_{c_{\sigma(2)}}^{c_2} \dots \delta_{c_{\sigma(n)}}^{c_n}, \quad (5)$$

where  $c_i$  is the color of the  $i$ th leg and  $\bar{c}_{\sigma(i)}$  is the anticolor of the line connected to the  $i$ th leg. The color indices  $c_i$  and  $\bar{c}_{\sigma(i)}$  take on values between 1 and  $N_C = 3$ . In the color flow basis, the color operators ( $\mathbf{T}$ ) can be decomposed as

$$\mathbf{T}_i = \lambda_i \mathbf{t}_{c_i} - \bar{\lambda}_i \bar{\mathbf{t}}_{\bar{c}_i} - \frac{1}{N} (\lambda_i - \bar{\lambda}_i) \mathbf{s}, \quad (6)$$

where  $\lambda_i$  and  $\bar{\lambda}_i$  are variables used to define the type of color object, and their values are given in Table I for quarks, antiquarks, and gluons<sup>2</sup>. The operators  $\mathbf{t}$ ,  $\bar{\mathbf{t}}$ , and  $\mathbf{s}$  are defined by their operation on a basis tensor.

The color flow basis is over-complete and not orthogonal. Therefore, when calculating the total color factor for a given amplitude, one needs to sum over all combinations of products of the basis tensors. For any given amplitude, the number of possible color flows is given by  $(n_q + n_g)!$ , where  $n_q$  is the number of external quark-antiquark pairs, and  $n_g$  is the number of external gluons. At first glance, this scaling suggests that the color flow basis is suboptimal to calculate color coefficients for many-parton states. This scaling is however outweighed by the fact that the color flow basis allows one to efficiently “build up” the color coefficient for a definite color structure with  $n$  partons from the coefficients of definite  $(n-1)$  parton color configurations<sup>3</sup>.

In order to use the color flow representation in a parton shower, the splitting kernels need to be recalculated in the color flow basis. The most straight forward case is the splitting of the nonet or singlet gluon into a quark and an antiquark pair. The color flow for this splitting is completely determined by color conservation, and is therefore identical to the traditional splitting kernel. Gluon emission from (anti)quarks in the FC parton shower should be

<sup>2</sup> A similar decomposition in the color-trace basis is discussed in Ref. [7].

<sup>3</sup> Using a similar method for orthogonal color bases [29] appears to be less straightforward.

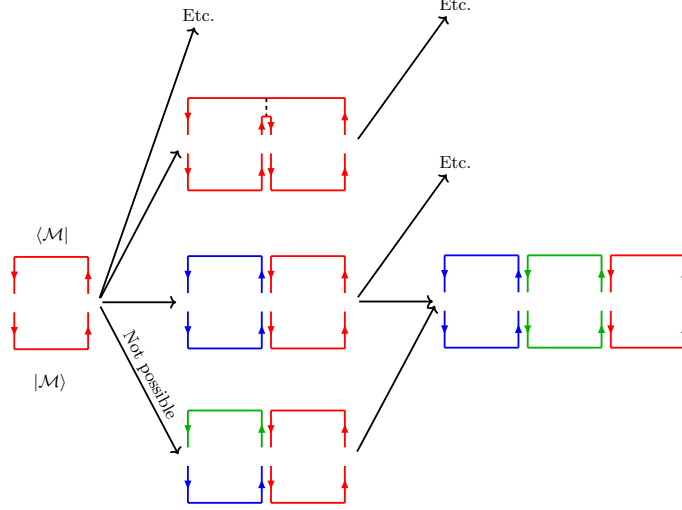


FIG. 1. Example for the sampling of color configurations. Note that both color configurations  $\langle \mathcal{M} |$  and  $| \mathcal{M} \rangle$  are kept throughout the evolution. A detailed discussion can be found in the main text.

governed by two splitting kernels – one to describe the emission of a nonet gluon ( $P_{qq}^{(9)}$ ), and another for singlet gluon emissions ( $P_{qq}^{(1)}$ ). Using Catani-Seymour dipole kernels [12] in the variables of the DIRE parton shower [13]

$$P_{qq} = C_F \left( \frac{2(1-z)}{(1-z)^2 + \kappa^2} - (1+z) \right), \quad P_{gg} = \frac{C_A}{2} \left( \frac{2(1-z)}{(1-z)^2 + \kappa^2} - 2 + z(1+z) \right), \quad P_{gq} = \frac{T_R}{2} (1 - 2z(1-z))$$

to capture the phase space dependence of the branching, the color flow specific kernels can be defined as

$$P_{qq}^{(9)}(x) = \lambda_i \mathbf{t}_{c_i} \frac{P_{qq}(x)}{C_F} \quad (7)$$

$$P_{qq}^{(1)}(x) = -\frac{\lambda_i}{N_C} \mathbf{s} \frac{P_{qq}(x)}{C_F}. \quad (8)$$

One splitting kernel is included for each possible change (due to the action of  $\mathbf{t}_{c_i}$ ) to the color structure, with some examples given in Fig 2. The branching of a gluon into two gluons can be distributed over two nonet gluon splitting kernels,

$$P_{gg}^{(+)} = \lambda_i \mathbf{t}_{c_i} \frac{P_{gg}(x)}{C_A}, \quad (9)$$

$$P_{gg}^{(-)} = -\bar{\lambda}_i \bar{\mathbf{t}}_{\bar{c}_i} \frac{P_{gg}(x)}{C_A}, \quad (10)$$

where  $P_{gg}^{(+)}(P_{gg}^{(-)})$  is the splitting off of the color(anticolor) gluon line, respectively. Examples are again shown in Fig 2. Note that in the color flow basis, nonet and singlet gluons do not couple directly. Hence, splitting functions involving both nonet and singlet gluons are absent.

## B. Implementation as a Parton Shower

The leading color approximation usually employed in parton showers removes all color interference effects at  $\mathcal{O}(1/N_C^2)$ . Naively, this would also occur when using the color-flow specific splitting functions described in the previous section, although some subleading contributions due to extended spectator assignments and the presence of singlet gluons would be retained.

The recovery of interference terms is central to defining a consistent FC parton shower. One possible solution is to sample color configurations stochastically, i.e. organize the evolution of the color configuration as a step-by-step splitting process from one definite color structure to another definite color structure. The correct inclusion of color factors then requires that there exists at least one path of splittings populating a definite color configuration, and

that the correct color-factor of the configuration can be calculated efficiently. This strategy is illustrated in Fig. 1. There, the  $\langle \mathcal{M} |$  state contains the emission from the spectator line, while the  $|\mathcal{M}\rangle$  state includes the emission from the radiator line. In leading color showers, both amplitudes are identical. The first step in moving to FC showers, is to remove this simplification. This is easily achieved for the first emission from a single color-anticolor dipole (i.e. with only one possible assignment of radiator and spectator) by employing an eikonal radiation pattern.

As already alluded to, complications arise in handling the states  $\langle \mathcal{M}' |^* \neq |\mathcal{M}\rangle$  after the first emission: to recover Eq. 2, the spectator assignment needs to be extended to include all partons. Interference effects can be handled by a Monte Carlo sum over states. For this, both  $\langle \mathcal{M}' |$  and  $|\mathcal{M}\rangle$  are kept throughout the evolution. For a change  $|\mathcal{M}\rangle \rightarrow |\tilde{\mathcal{M}}\rangle$  that was chosen with the Sudakov veto algorithm, a definite  $\langle \tilde{\mathcal{M}}' |$  is chosen based on all possible non-zero probabilities  $\langle \tilde{\mathcal{M}}' | \tilde{\mathcal{M}} \rangle$ . Both  $\langle \tilde{\mathcal{M}} |$  and  $|\tilde{\mathcal{M}}'\rangle$  are then used in the next evolution step when probing a subsequent color configuration. This is permissible if we choose to keep the phase-space dependence of the splitting kernels independent of the color configuration  $\langle \mathcal{M}' |$ .

A technical complication that arises when adding in the interference effects in this Monte-Carlo fashion is that emissions from the radiator and emissions from the spectator are not necessarily associated with the same momentum assignments. This necessarily deteriorates the efficiency of the Monte-Carlo integration over phase space, since the phase-space structure probed by emissions from the radiator might approach different singular phase space points than the structure probed by emissions from the spectator in the  $k_T \rightarrow 0$  limit. To counter this issue, a cutoff  $t_{FC}^{\text{cut}}$  to transition from a FC shower to an LC shower is implemented. If the algorithm decides to admit an emission below this cutoff, then only the leading color connected parton can be the spectator, and the LC color factor is used. While this does introduce some error in the calculation, the cutoff can be pushed sufficiently low to have a negligible effect on the distributions as will be shown in Section IV. The main benefit of this cutoff is that the weight fluctuations are reduced and the convergence of the calculation is improved, requiring fewer events and less time to obtain smooth distributions.

With the issues discussed above in mind, the FC parton shower algorithm is described in detail below. The algorithm uses two accept-reject steps, along with a weighting factor to correct for the sign of the color factor. The first step in the algorithm involves a modification to the allowed colors for partons. Typically, for a leading color shower, the partons are labeled with a color and an anticolor index ranging from 1 to  $\infty$ . In the FC parton shower, color and anticolor indices range from 1 to 3 (i.e.  $r$ ,  $g$ , and  $b$ ). This begins with the labeling of the partons that are produced in the hard process. Multiple partons can have the same color index. This is important when calculating the color factor, and allows the calculation to be performed more efficiently.

The first accept-reject step is similar to a leading color shower, but with the extension that non-leading color connected partons are allowed to be spectators. This can be viewed as determining if a splitting is kinematically preferred: the accept-reject step determines the phase space variables  $(t, z)$ , which parton is considered the radiator and which the spectator. The phase space region  $t < t_{FC}^{\text{cut}}$  will be considered at leading color only. In this case, no further steps related to picking the color structure are necessary. Splittings with  $t > t_{FC}^{\text{cut}}$  will be considered with complete  $N_C = 3$  color factors, and the color factor needs to be corrected to obtain the full color result, if the splitting is not a gluon into a quark-antiquark pair. This is facilitated by calculating the color correlator

$$\langle \mathcal{M}' | \frac{\mathbf{T}_{\tilde{k}} \cdot \mathbf{T}_{\tilde{i}j}}{\mathbf{T}_{\tilde{i}j}^2} | \mathcal{M} \rangle, \quad (11)$$

where the  $\mathbf{T}$ 's are given by Eq. 6, and depends on the type of partons  $k$  and  $ij$ . This factor is calculated exactly, given a fixed radiator and spectator. This result is then used as a second accept-reject probability to correct to splitting kernel from the LC overestimate (produced by the first accept-reject step) to the correct overall color factor for the given radiator and spectator. Since the accept probability is not bounded by zero and one, we employ weighted parton shower techniques [21–23] in the second accept-reject step to ensure that the correct overall color factor is consistently exponentiated<sup>4</sup>. At this point, we have achieved the exponentiation of Eq 2. To allow iteration, it is necessary to fix a definite color structure for subsequent evolution. This will be discussed in the next section.

---

<sup>4</sup> Specifically, splittings will be accepted with probability  $f/g$ , and introduce corrective event weights. If the event is accepted, it receives the weight  $\frac{g}{h}$ . Rejection leads to an event weight  $\frac{g}{h} \frac{h-f}{g-f}$ . Here,  $f$ ,  $g$ , and  $h$  are given by

$$f = -\langle \mathcal{M}' | \mathbf{T}_k \cdot \mathbf{T}_{ij} | \mathcal{M} \rangle, \quad h = \langle \mathcal{M}' | \mathbf{T}_{ij}^2 | \mathcal{M} \rangle, \quad g = \begin{cases} -h & \frac{f}{g} < 0 \\ 2f & \frac{f}{g} > 1 \\ h & \text{otherwise} \end{cases}.$$

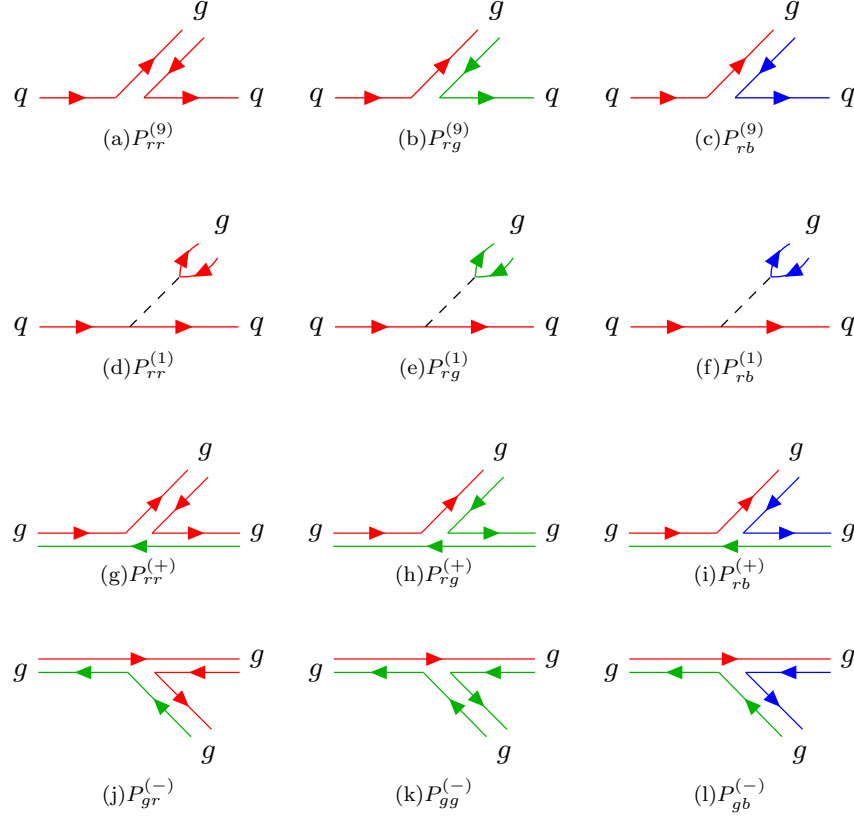


FIG. 2. Examples of possible splittings in the fixed color shower. The upper two rows show all possible splittings for a red quark, while the lower two rows show all splittings of a red-antigreen gluon. Note that in all cases, we implement one kernel for soft-tagged emission and one kernel for soft-tagged radiator.

### C. Color Flow Sampling

The overall color factor exponentiated into no-emission probabilities now correctly recovers the full color correlator (Eq. 2). To allow the iteration of the stochastic color sampling algorithm once a branching has been accepted, it is necessary to choose a definite color structure for the branching. This choice has to be commensurate with the contribution of the particular color configuration to the overall color factor. To obtain the correct result in the color flow basis, the color operators can be separated into different components, depending on if the splitting is  $P_{qq}^{(9)}$ ,  $P_{qq}^{(1)}$ ,  $P_{qq}^{(9)}$ ,  $P_{qq}^{(1)}$ ,  $P_{gg}^{(+)}$ , or  $P_{gg}^{(-)}$ . Six possible color structures for the radiator and six possible color structures for the spectator are allowed for given emitter  $ij$  with color  $c_i$  and spectator  $k$  with anticolor  $\bar{c}_k$ , giving in total 36 possible configurations. Possible splittings of a red quark and a red-antigreen gluon are illustrated in Fig. 2. An additional event weight needs to be applied to correctly account for picking one definite (out of 36 possible) color structures.

The correction is obtained by first calculating the “color weight” for each one of the 36 possible configurations,

$$P_{\alpha\beta} = |\langle \mathcal{M}' | \mathbf{t}_k^\alpha \cdot \mathbf{t}_{ij}^\beta | \mathcal{M} \rangle|, \quad (12)$$

where  $\mathbf{t}_k^\alpha$  is one of the six possible color factors for the color insertion on the spectator line acting on  $\langle \mathcal{M}' |$ , and  $\mathbf{t}_{ij}^\beta$  is color insertion on the radiator line acting on  $|\mathcal{M}\rangle$ . The absolute value ensures a positive definite probability to pick one color structure,

$$P = \frac{P_{\alpha\beta}}{\sum_{\alpha,\beta} P_{\alpha\beta}}. \quad (13)$$

The color structure can then be chosen probabilistically according to Eq. 13. Finally, to obtain the correct weight for the branching, the sign of this contribution to the overall color factor needs to be reinstated. Hence, an additional

event weight

$$\frac{g_{col}}{h_{col}} = \frac{\mathbf{t}_k^\alpha \cdot \mathbf{t}_{ij}^\beta}{|\mathbf{t}_k^\alpha \cdot \mathbf{t}_{ij}^\beta|} \frac{\sum_{\alpha,\beta} |\mathbf{t}_k^\alpha \cdot \mathbf{t}_{ij}^\beta|}{\sum_{\alpha,\beta} \mathbf{t}_k^\alpha \cdot \mathbf{t}_{ij}^\beta}. \quad (14)$$

is applied. At this point, the emission is generated according to Eq. 2, and a definite color configuration is chosen to allow iteration of the stochastic color sampling. The real-emission kinematics are constructed and, and the algorithm returns to the beginning again.

If a gluon decays into a quark-antiquark pair, the color factor is determined explicitly by color conservation. Hence, once a proposed branching passes the first accept-reject step, the kinematics of the quark and antiquark are generated and the next emission is considered.

The complete algorithm is summarized, for convenience, as a flow chart in Fig. 8. An example of the first two branchings for the evolution of  $\langle \mathcal{M} |$  and  $|\mathcal{M} \rangle$  is shown in Fig. 1. In this example, we begin with a red quark and an antired antiquark, and calculate all 36 possible color configurations after one emission. Only three of the 36 are shown explicitly in the example. The probabilities of ending up in each of the three sample states shown are, from the top down,  $\frac{1}{12}$ ,  $\frac{1}{4}$  and zero. The last contribution is vanishing, because the inner product for the given two color flows is exactly zero. However, the algorithm still explores all of color space, even if one history to reach the desired state has zero probability – as indicated by the central row. As long as any history with non-vanishing probability exists, the color configuration is sampled, including the correct color factor.

#### D. Further technicalities

The FC parton shower algorithm relies on weighted parton shower techniques to ensure correct exponentiation of the color correlator. As such, it is possibly prone to numerical instabilities because of the accumulation of large event weights if  $\langle \mathcal{M}' | \mathbf{T}_{ij}^2 | \mathcal{M} \rangle$  is a very poor overestimate of  $-\langle \mathcal{M}' | \mathbf{T}_k \cdot \mathbf{T}_{ij} | \mathcal{M} \rangle$ . If this is repeatedly the case, event weights can become large, thus requiring many more events to appropriately obtain the true result. This can for example be the case if high-multiplicity configurations are probed very often at small ordering parameter. Therefore, to mitigate this effect, we introduce a color sampling cutoff scale ( $t_{FC}^{\text{cut}}$ ). Below this scale, we match the FC parton shower to a leading color (LC) parton shower. Note that we do not restrict the number of emissions that carry the correct color factor. Matching the FC parton shower to leading color using only a resolution scale defines an infrared safe method, whereas infrared safety would have to be re-evaluated if the FC parton shower were terminated after a fixed number of emissions.

Since both the LC parton shower and the FC parton shower rely on the color flow basis, the transition to leading color is straightforward: an auxiliary color structure with leading color connected partners is tracked during FC evolution. The only complication that arises is in handling the color singlets. In this case, mapping onto a leading color configuration means that the emitting parton keeps the original leading color connected partner, and the singlet is only leading color connected to itself. Additionally, singlet gluons in the auxiliary leading color configuration will only be allowed branch into quark-antiquark pairs once the LC parton shower evolves the state. This follows since gluon radiation from the color line is exactly canceled by gluon radiation from the anticolor line. It is beneficial to enforce this cancellation exactly, rather than achieving it stochastically by averaging over events.

Once the parton shower evolution is terminated at the overall cutoff  $t^{\text{cut}}$ , the event transitions to subsequent hadronization. In this paper, we employ the Lund string model [18] as implemented in PYTHIA 8.2 [19, 20] for hadronization. This model is constructed in the  $N_C \rightarrow \infty$  limit, and thus not immediately suitable for projecting  $N_C = 3$  partonic states onto a spectrum of hadrons. However, note that the  $N_C \rightarrow \infty$  approximation is less ad-hoc since the FC parton shower is not used all the way down to  $t^{\text{cut}} < t_{FC}^{\text{cut}}$ , meaning that a perturbative mapping to a leading color configuration has already been obtained. Beyond this mapping, it is important to clarify how  $N_C = 3$  color configurations are prepared for hadronization. This in particular concerns the treatment of gluons.

Within the Lund string model, gluons are interpreted as links in a chain of color dipoles, i.e. a color string. These links only induce transverse “kinks” on the string, as further parameters would impede an infrared safe matching onto partonic states. When preparing a fixed color configuration for hadronization, we need to ensure that this quality is preserved. Otherwise, perturbative cancellations between color singlet and nonet gluon states are distorted and potentially invalidated, and the result is not infrared safe. To avoid these issues, a second auxiliary leading color configuration is kept throughout the (FC and LC) evolution. In this configuration, all emissions of a singlet gluon are replaced by a nonet gluon, retaining the kinematics of (FC or LC) event. This event is not used to calculate color factors, and is only used as input for hadronization, instead of the final event after the parton shower has terminated. This guarantees that hadronization effects do not spoil the cancellation required to remove the trace component of nonet gluons by means of singlet gluons.

## IV. RESULTS

We have as implemented the FC parton shower as PYTHON code, building on the implementation in [17]. The code is interfaced to PYTHIA through Les Houches event files [30]. This section presents results of the new shower algorithm for  $e^+e^- \rightarrow$  jets scattering events, split into a validation section and some preliminary comparisons to LEP data. All results have been generated with an overall parton shower cutoff  $t^{\text{cut}} = 1.0$  GeV and a fixed color cutoff of  $t_{FC}^{\text{cut}} = 3$  GeV. The LC baseline used as a comparison is a direct implementation of the final-state shower presented in [13]. This baseline uses  $C_F = \frac{4}{3}$  for gluon emissions from (anti)quark lines, and explicitly includes two splitting kernels for the two possible color-assignments for the  $g \rightarrow gg$  branching.

### A. Validation

This section illustrates the differences of FC and LC parton showers. To gain insight into these differences, we use observables that should be particularly sensitive. The impact of color corrections can be significant when examining states of fixed parton multiplicity, e.g. after the second parton shower emission. Instead, the observables used in this section will be based on jets to ensure a realistic assessment of the impact.

At the level of the second emission, FC and LC parton shower start deviating, because color factors become non-trivial, and because non-leading color radiator-spectator dipoles appear. Color corrections should thus influence the distribution of four-jet events, in particular the differences of dijet invariant masses in four-jet events. Jet masses are heavily affected by non-perturbative corrections. This suggests testing the impact of color corrections on the dijet mass ratio variables  $r_{ij,kl} = \frac{s_{ij}}{s_{kl}}$ , where  $s_{ij} = (p_i + p_j)^2$  and  $p_i$  are energy-ordered jet momenta. The two hardest jets (index 3 and 4) should be a reasonable proxy for regions enriched by the primary quark/antiquark. Secondary emissions from the quark-antiquark dipole are allowed in the FC parton shower, but absent in the LC version. We thus expect ratios containing  $s_{34}$  to exhibit color corrections. A similar argument holds for the “eikonal pattern”  $\kappa_{364} = \frac{s_{36}s_{64}}{s_{34}s_{3456}}$ , which probes the radiation from the quark-antiquark dipole.

Representative results are shown in Fig 3. As expected, the FC parton shower yields a reduced  $s_{34}$ , thereby producing harder  $r_{46,34}$  and  $r_{56,34}$  spectra in Figs. 3(a) and 3(b), respectively. The eikonal pattern  $\kappa_{364}$  (Fig. 3(c)) also shows the impact of the FC parton shower. Figures 3(d)-3(c) however accentuate that the impact of hadronization is still non-negligible. At hadron level, the color corrections of  $\mathcal{O}(5\%)$  are found.

### B. Comparison to data

This section compares the LC and FC parton shower to selected LEP data. Since the FC parton shower directly builds upon a PYTHON implementation for Dire, we use the default tune for the DIRE plugin to PYTHIA (see App. A) to set non-perturbative parameters. This tune was performed with a version of DIRE that includes  $\mathcal{O}(\alpha_s^2)$  corrections designed to recover the inclusive NLO corrections to soft-gluon emissions, i.e. the two-loop cusp anomalous dimension [34–37]. In the current study, we do not include these corrections, since the two loop cusp anomalous dimension partially arises from sub-leading color effects which are already captured by the FC parton shower. To avoid double-counting, and to keep the comparison on even footing, we therefore do not include such corrections in the (LC or FC) parton shower. This deteriorates the comparison to LEP data, so we use  $\alpha_s(M_Z) = 0.125$  for a more sensible baseline. This discussion already emphasizes that the quality of the data comparisons should not be considered final before the event generator has been retuned. Such a procedure will however obscure the differences between LC and FC shower, and is thus avoided here.

Figure 4 compares the LC and FC parton showers to ALEPH data [38]. As anticipated, the parton showers are not an ideal representation of the data. The FC shows a reduced rate of semi-hard emissions compared to LC, which also induces a reduced Sudakov suppression at small thrust. A similar trend can also be observed in the related jet broadening shown in Fig. 5. This reduction of overall activity can be traced to negative interference contributions due to the presence of color singlet gluons. The overall effect of subleading contributions is, however, moderate.

Larger effects are anticipated for differential clustering scales. Figure 6 compares the jet separation between two- and three-jet events to OPAL measurements [39]. The result of LC and FC parton showers are very similar, since the observable is dominated by the first parton shower emission, which is identical in both algorithms. A more drastic change can be found in the jet separation between three- and four-jet states  $y_{34}$  shown in Fig. 7. The impact of the modified kinematics because of radiation from sub-leading color dipoles produces a much softer spectrum, as already observed to a lesser degree in Figs. 4 and 5. Of course, it needs to be noted that small differences between LC and FC showers in the non-perturbative region ( $y_{34} \lesssim 5 \cdot 10^{-4}$ ) can have significant impact of  $\mathcal{O}(10\%)$  also in the hard region because of the normalization of the plot. In general, we do however observe a tendency towards softer multi-jet



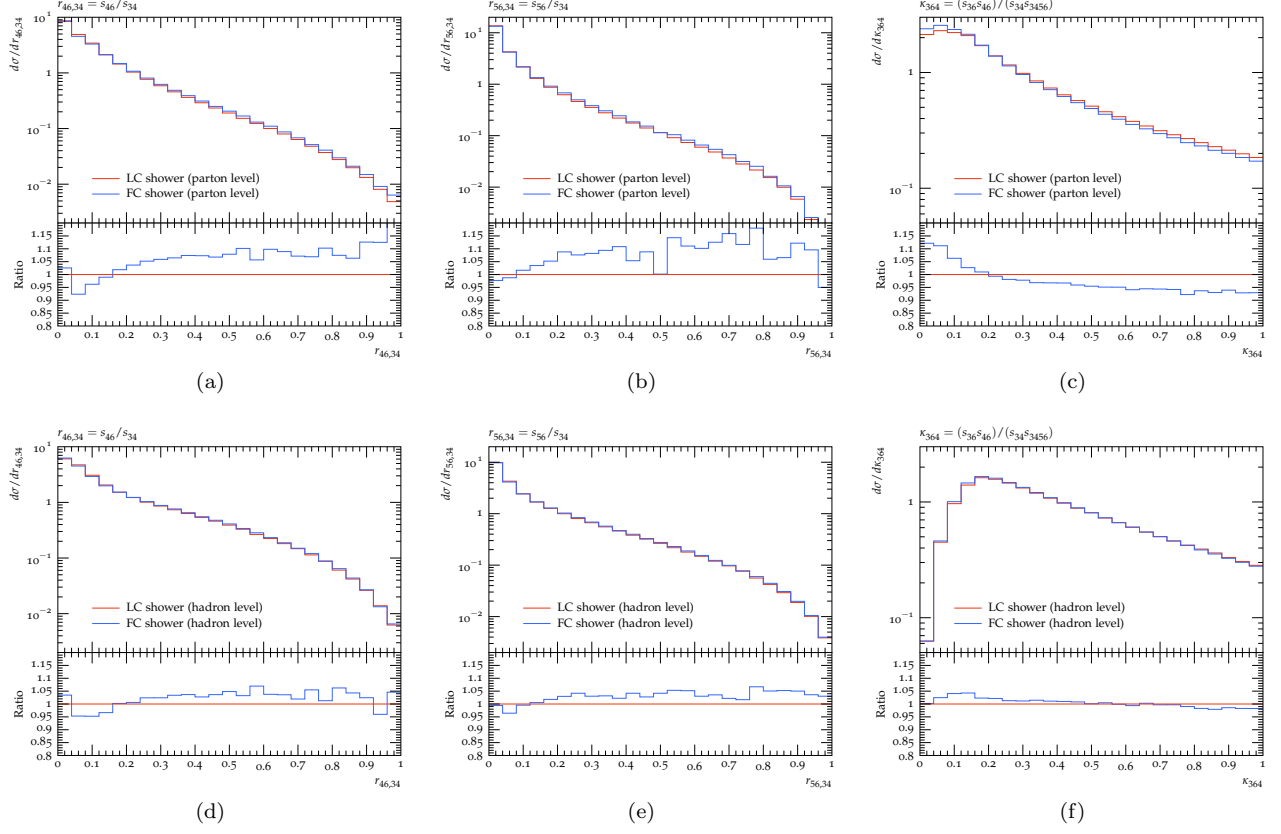


FIG. 3. Comparison of the LC parton shower to the FC parton shower. The definition of the dijet mass ratios  $r_{ij,kl}$  and the scaled jet eikonal can be found in the main text. The upper row gives parton level results, while the lower row shows results at hadron level. are defined using the Durham algorithm [31] as implemented in the textscFastjet package [32], clustering to exactly four jets with  $p_T > 3$  GeV. All figures have been produced using RIVET [33].

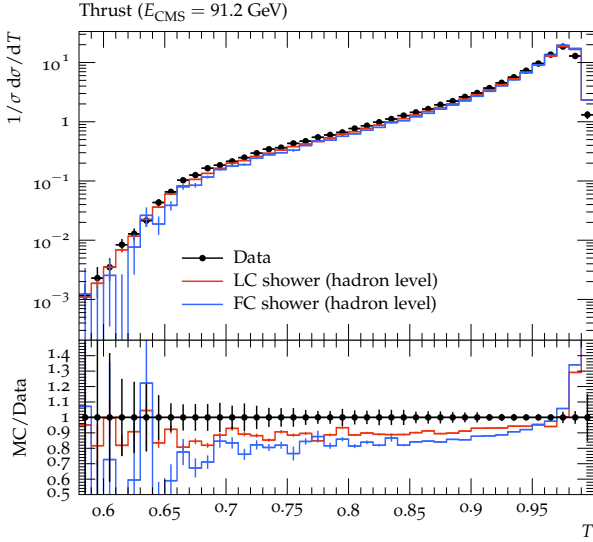


FIG. 4. Thrust as measured by ALEPH [38] and implemented in RIVET [33], compared to leading color and fixed color parton showers.

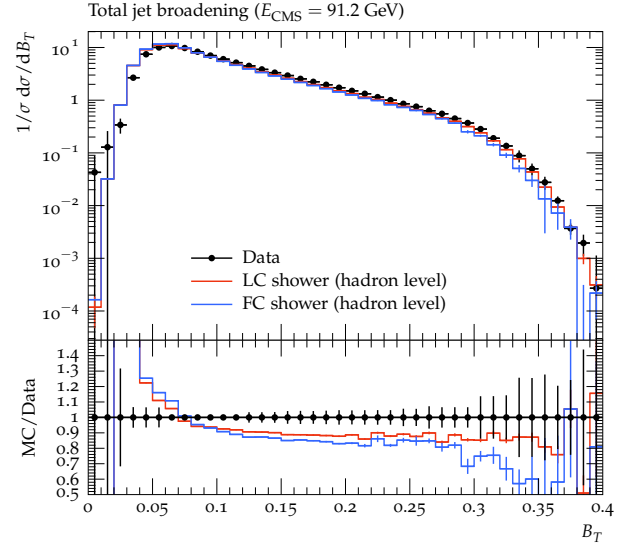


FIG. 5. Total jet broadening as measured by ALEPH [38] and implemented in RIVET [33], compared to leading color and fixed color parton showers.

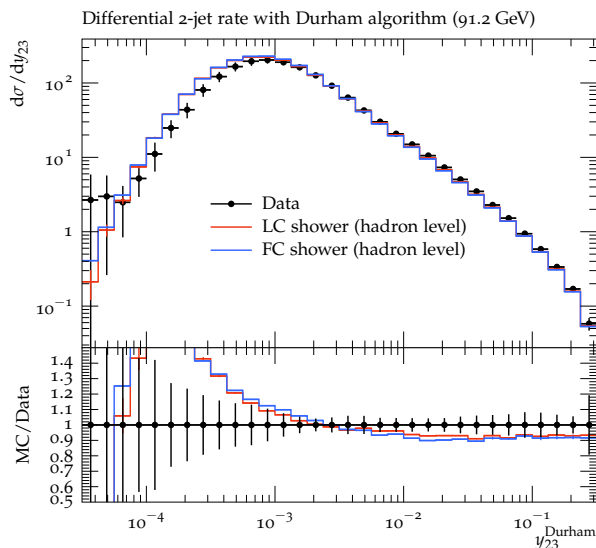


FIG. 6. Jet separation between two- and three-jet configurations in the Durham algorithm, as measured by OPAL [39] and implemented in RIVET [33], compared to leading color and fixed color parton showers.

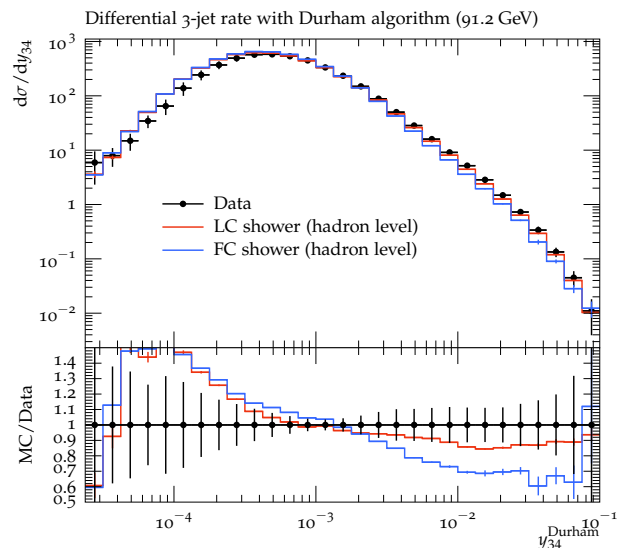


FIG. 7. Jet separation between three- and four-jet configurations in the Durham algorithm, as measured by OPAL [39] and implemented in RIVET [33], compared to leading color and fixed color parton showers.

spectra in the FC parton shower. This is expected because of the inclusion of depleting sub-leading corrections. The kinematics of these corrections is, in our algorithm, completely determined by the splitting  $|\mathcal{M}\rangle$ , and not influenced by the potentially different propagator structure of splittings in  $\langle\mathcal{M}|$ . The data comparison raises the question if subleading color corrections can really be treated independently from other, kinematic, corrections. We believe that this study gives important input for future developments of more accurate parton showers.

## V. CONCLUSIONS AND OUTLOOK

This article has presented the first implementation of a fixed color parton shower algorithm that remains numerically feasible and stable for an arbitrary number of emissions. This has been achieved by leveraging the qualities of the color flow basis. The growth in complexity has been tamed by employing a stochastic sampling of color configurations. Remaining numerical instabilities (that could be ameliorated by accumulating higher statistics) have been improved with the introduction of a cutoff on fixed color evolution. This necessitates keeping track of auxiliary events that allow a matching onto a leading color shower. We have also discussed an infrared safe matching to the Lund string hadronization model.

Preliminary comparisons to LEP data have been presented. Here, the data description should not be regarded as final – in particular because known higher-order corrections are not present to avoid double-counting in the fixed color result – but rather as allowing to assess benefits of differences between leading- and fixed color evolution. These comparisons indicate an unfavorable trend towards too soft radiation patterns in the fixed color parton shower. This observation is an important input for efforts to define parton showers at higher accuracy, since it suggests that the exponentiation of complete color correlators without also including multi-parton kinematic correlations (beyond the three-particle correlations given by Eq. 2) might have undesirable consequences.

The algorithm has been implemented as stand-alone PYTHON code that was interfaced to PYTHIA via Les Houches files. A natural next step is to extend the fixed color evolution also to initial-state splittings. Due to the handling of multiple parton interactions and beam remnants, this will require a more native implementation within the event generator framework. Further logical next steps would be the introduction of  $\mathcal{O}(\alpha_s^2)$  corrections to allow recovering the two-loop cusp-anomalous dimension, and the inclusion of multi-parton kinematic correlations. In the future it will be important to study the complex interplay of sub-leading color, kinematics, and higher order corrections to better describe the data.

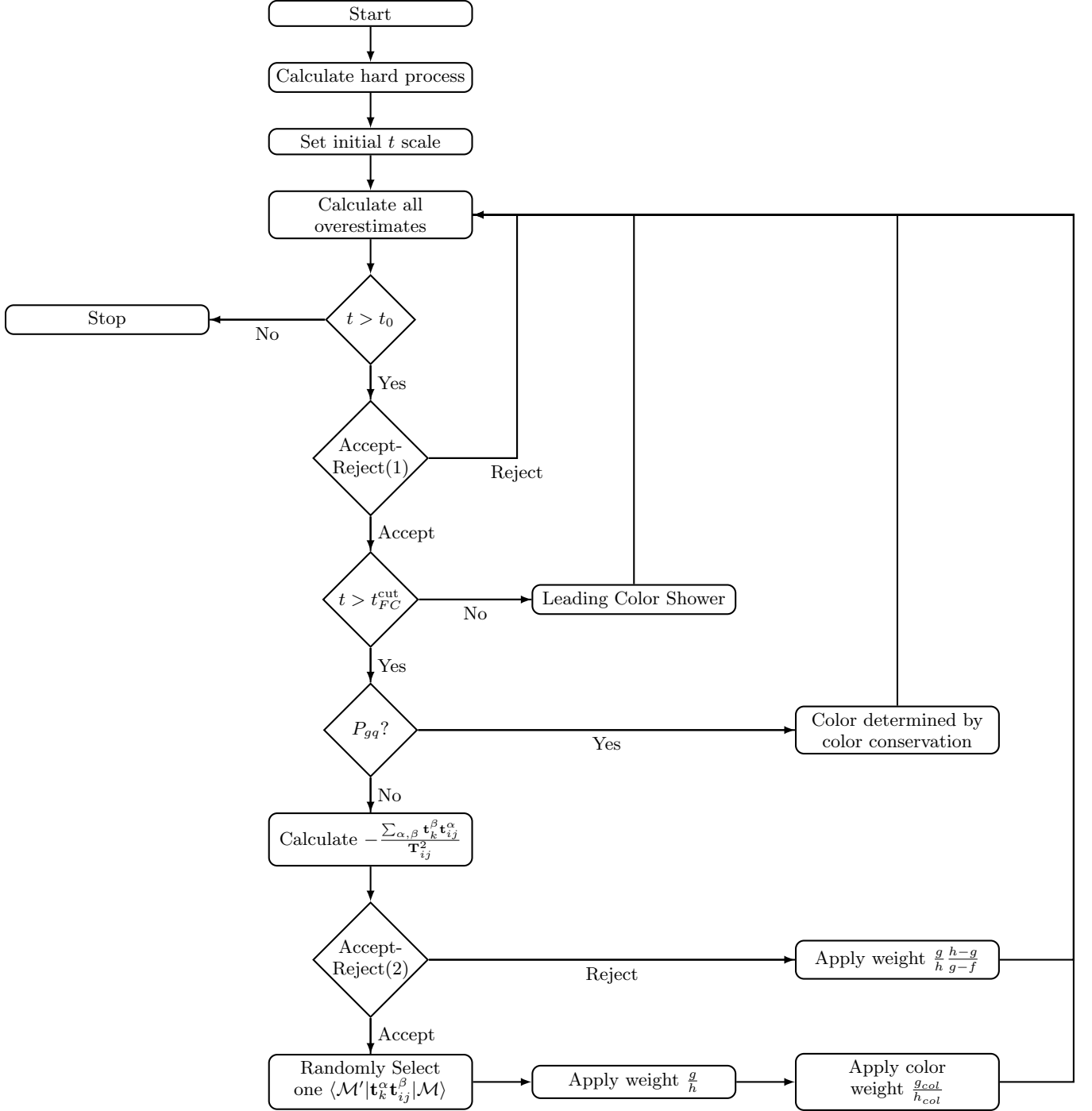


FIG. 8. Flowchart representing the FC parton shower algorithm. Details on the calculation and of the definitions of  $f$ ,  $g$ ,  $h$ ,  $g_{col}$ , and  $h_{col}$  can be found in Sec. III B and Sec. III C.

### ACKNOWLEDGMENTS

We thank S. Höche for help in developing the algorithm described in Sec. III through numerous discussions, and for pointing out the option to use an independent cutoff for fixed color evolution. We further thank J. Forshaw and S. Plätzer for interesting discussions on the Glauber gluons, and S. Keppeler and M. Sjödaahl for discussions on color bases. This work was supported by Fermi Research Alliance, LLC, under Contract No. DE-AC02-07CH11359 with the U.S. Department of Energy, Office of Science, Office of High Energy Physics.

## Appendix A: The default tune of the Dire plugin for Pythia

Since the default tune of the DIRE plugin to PYTHIA has not been documented beyond comments in the `direforpythia` code repository and source code, we here give the tune settings, as introduced in revision SHA `ca99627ea3c6741ddb98ea55054c44c6b298c0ba` of [gitlab.com/dire/direforpythia](https://gitlab.com/dire/direforpythia):

```
StringPT:sigma          = 0.2952
StringZ:aLund           = 0.9704
StringZ:bLund           = 1.0809
StringZ:aExtraDiquark   = 1.3490
StringFlav:probStoUD    = 0.2046
StringZ:rFactB          = 0.8321
StringZ:aExtraSQuark    = 0.0
TimeShower:pTmin        = 0.9
```

These tune settings have been extracted by comparing against  $e^+e^- \rightarrow \text{hadrons}$  data from  $\sqrt{s} = 14 - 91.2$  GeV, as well as comparisons against Tevatron and LHC data. The following DIRE plugin-specific settings were assumed

```
PDF:pSet                = LHAPDF6:MMHT2014nlo68cl
PDF:pHardSet            = LHAPDF6:MMHT2014nlo68cl
TimeShower:alphaSvalue  = 0.1201
SpaceShower:alphaSvalue = 0.1201
ShowerPDF:usePDFalphas  = on
ShowerPDF:useSummedPDF  = on
DireSpace:forceMassiveMap = on
ShowerPDF:usePDFmasses  = off
DireTimes:kernelOrder   = 1
DireSpace:kernelOrder   = 1
```

This implies that the soft-collinear pieces of splitting kernels were rescaled with corrections to recover the inclusive  $\mathcal{O}(\alpha_s^2)$  correction in the soft limit (see e.g. [40]).

- 
- [1] A. Buckley *et al.*, Phys. Rept. **504**, 145 (2011), arXiv:1101.2599 [hep-ph].
  - [2] G. C. Fox and S. Wolfram, Nucl. Phys. **B168**, 285 (1980); T. Sjöstrand, Phys. Lett. **B157**, 321 (1985); B. Webber, Ann. Rev. Nucl. Part. Sci. **36**, 253 (1986); G. Gustafson and U. Pettersson, Nucl. Phys. **B306**, 746 (1988).
  - [3] V. N. Gribov and L. N. Lipatov, Sov. J. Nucl. Phys. **15**, 438 (1972); Y. L. Dokshitzer, Sov. Phys. JETP **46**, 641 (1977); G. Altarelli and G. Parisi, Nucl. Phys. **B126**, 298 (1977).
  - [4] S. Plätzer and M. Sjödaahl, JHEP **07**, 042 (2012), arXiv:1201.0260 [hep-ph].
  - [5] J. Bellm, (2018), arXiv:1801.06113 [hep-ph].
  - [6] Z. Nagy and D. E. Soper, JHEP **07**, 119 (2015), arXiv:1501.00778 [hep-ph].
  - [7] Z. Nagy and D. E. Soper, JHEP **09**, 114 (2007), arXiv:0706.0017 [hep-ph].
  - [8] Z. Nagy and D. E. Soper, JHEP **1206**, 044 (2012), arXiv:1202.4496 [hep-ph].
  - [9] R. Á. Martínez, M. De Angelis, J. R. Forshaw, S. Plätzer, and M. H. Seymour, (2018), arXiv:1802.08531 [hep-ph].
  - [10] Z. Nagy and D. E. Soper, JHEP **10**, 019 (2016), arXiv:1605.05845 [hep-ph]; (2017), arXiv:1711.02369 [hep-ph].
  - [11] Y. I. Azimov, Y. L. Dokshitzer, V. A. Khoze, and S. Troyan, Z.Phys. **C27**, 65 (1985); Y. I. Azimov, Y. L. Dokshitzer, V. A. Khoze, and S. I. Troian, Phys. Lett. **165B**, 147 (1985); L. Lönnblad, Comput. Phys. Commun. **71**, 15 (1992).
  - [12] S. Catani and M. H. Seymour, Nucl. Phys. **B485**, 291 (1997), hep-ph/9605323; S. Schumann and F. Krauss, JHEP **03**, 038 (2008), arXiv:0709.1027 [hep-ph].
  - [13] S. Höche and S. Prestel, Eur. Phys. J. **C75**, 461 (2015), arXiv:1506.05057 [hep-ph].
  - [14] B. R. Webber, Nucl. Phys. **B238**, 492 (1984); G. Marchesini and B. R. Webber, Nucl. Phys. **B238**, 1 (1984); Nucl. Phys. **B310**, 461 (1988); Z. Nagy and D. E. Soper, JHEP **06**, 178 (2014), arXiv:1401.6366 [hep-ph]; S. Plätzer and S. Gieseke, JHEP **01**, 024 (2011), arXiv:0909.5593 [hep-ph]; W. T. Giele, D. A. Kosower, and P. Z. Skands, Phys. Rev. **D84**, 054003 (2011), arXiv:1102.2126 [hep-ph].
  - [15] F. Dulat, S. Höche, and S. Prestel, (2018), arXiv:1805.03757 [hep-ph].
  - [16] R. Angeles-Martinez, J. R. Forshaw, and M. H. Seymour, JHEP **12**, 091 (2015), arXiv:1510.07998 [hep-ph].
  - [17] S. Höche, Tutorial on Parton Showers, CTEQ Summer School 2015, unpublished.
  - [18] B. Andersson, G. Gustafson, G. Ingelman, and T. Sjöstrand, Phys. Rept. **97**, 31 (1983).
  - [19] T. Sjostrand, S. Mrenna, and P. Z. Skands, JHEP **05**, 026 (2006), arXiv:hep-ph/0603175 [hep-ph].

- [20] T. Sjöstrand, S. Ask, J. R. Christiansen, R. Corke, N. Desai, P. Ilten, S. Mrenna, S. Prestel, C. O. Rasmussen, and P. Z. Skands, *Comput. Phys. Commun.* **191**, 159 (2015), arXiv:1410.3012 [hep-ph].
- [21] S. Höche, S. Schumann, and F. Siegert, *Phys. Rev.* **D81**, 034026 (2010), arXiv:0912.3501 [hep-ph].
- [22] S. Plätzer and M. Sjö Dahl, *Eur. Phys. J. Plus* **127**, 26 (2012), arXiv:1108.6180 [hep-ph].
- [23] L. Lönnblad, *Eur.Phys.J.* **C73**, 2350 (2013), arXiv:1211.7204 [hep-ph].
- [24] J. C. Collins and G. F. Sterman, *Nucl. Phys.* **B185**, 172 (1981).
- [25] G. T. Bodwin, S. J. Brodsky, and G. P. Lepage, *Phys. Rev. Lett.* **47**, 1799 (1981).
- [26] R. Angeles Martinez, J. R. Forshaw, and M. H. Seymour, *Phys. Rev. Lett.* **116**, 212003 (2016), arXiv:1602.00623 [hep-ph].
- [27] L. Lonnblad, *Z. Phys.* **C70**, 107 (1996); E. Avsar, G. Gustafson, and L. Lonnblad, *JHEP* **01**, 012 (2007), arXiv:hep-ph/0610157 [hep-ph]; C. Bierlich, G. Gustafson, L. Lönnblad, and A. Tarasov, *JHEP* **03**, 148 (2015), arXiv:1412.6259 [hep-ph].
- [28] F. Maltoni, K. Paul, T. Stelzer, and S. Willenbrock, *Phys. Rev.* **D67**, 014026 (2003), hep-ph/0209271.
- [29] S. Keppeler and M. Sjö Dahl, *JHEP* **09**, 124 (2012), arXiv:1207.0609 [hep-ph].
- [30] J. Alwall *et al.*, *Monte Carlos for the LHC: A Workshop on the Tools for LHC Event Simulation (MC4LHC) Geneva, Switzerland, July 17-16, 2006*, *Comput. Phys. Commun.* **176**, 300 (2007), arXiv:hep-ph/0609017 [hep-ph].
- [31] S. Catani, Y. L. Dokshitzer, M. Olsson, G. Turnock, and B. R. Webber, *Phys. Lett.* **B269**, 432 (1991).
- [32] M. Cacciari, G. P. Salam, and G. Soyez, *Eur. Phys. J.* **C72**, 1896 (2012), arXiv:1111.6097 [hep-ph].
- [33] A. Buckley, J. Butterworth, L. Lonnblad, D. Grellscheid, H. Hoeth, J. Monk, H. Schulz, and F. Siegert, *Comput. Phys. Commun.* **184**, 2803 (2013), arXiv:1003.0694 [hep-ph].
- [34] J. Kodaira and L. Trentadue, *Phys.Lett.* **B112**, 66 (1982).
- [35] C. Davies and W. J. Stirling, *Nucl.Phys.* **B244**, 337 (1984).
- [36] C. Davies, B. Webber, and W. Stirling, *Nucl. Phys.* **B256**, 413 (1985).
- [37] S. Catani, E. D’Emilio, and L. Trentadue, *Phys.Lett.* **B211**, 335 (1988).
- [38] A. Heister *et al.* (ALEPH), *Eur. Phys. J.* **C35**, 457 (2004).
- [39] P. Pfeifenschneider *et al.* (JADE, OPAL), *Eur. Phys. J.* **C17**, 19 (2000), arXiv:hep-ex/0001055 [hep-ex].
- [40] S. Höche, S. Mrenna, S. Prestel, M. Schönherr, and P. Z. Skands, in *10th Les Houches Workshop on Physics at TeV Colliders (PhysTeV 2017) Les Houches, France, June 5-23, 2017* (2018) arXiv:1803.07977 [hep-ph].